Econometric Analysis

Ch.3 Basic Asymptotic Theory

Ryuichi Tanaka

Theoretical Concerns in Econometrics

- Q1: Is your estimator consistent? (minimum requirement. Prove using the weak law of large number)
- Q2: What is the asymptotic distribution of your estimator? (needed for tests. Prove using the central limit theorem)

- We need to know the property of estimators when the sample size goes to infinity (asymptotic analysis)
- What does the convergence mean? (Convergence in probability, convergence in distribution)

3.1 Convergence of Deterministic Sequences

Definition 3.1:

(1) A sequence of nonrandom numbers $\{a_N : N = 1,2,\cdots\}$ converges to a (has limit a) if for all $\varepsilon > 0$, there exists N_ε such that if $N > N_\varepsilon$ then $|a_N - a| < \varepsilon$.

(2) A sequence $\{a_N : N = 1,2,\cdots\}$ is bounded if and only if there is some $b < \infty$ such that $|a_N| < b$ for all $N = 1,2,\cdots$.

Example 3.1

(1) $a_N = 2 + 1/N \to 2$ and bounded.

(2) $a_N = (-1)^N$ has no limit but is bounded.

(3) $a_N = N^{\frac{1}{4}}$ is not bounded (unbounded).

Large O and small o

Definition 3.2:

(1) A sequence $\{a_N\}$ is $O(N^\lambda)$ (at most of order N^λ) if $N^{-\lambda}a_N$ is bounded. When $\lambda = 0$, $\{a_N\}$ is bounded, and we also write $a_N = O(1)$.

(2) $\{a_N\}$ is $o(N^\lambda)$ if $N^{-\lambda}a_N \to 0$. When $\lambda = 0$, $\{a_N\}$ converges to zero, and we also write $a_N = o(1)$.

Note: $\{a_N\} = o(N^\lambda) \Rightarrow \{a_N\} = O(N^\lambda)$

Example 3.2:

(1) $a_N = \log(N)$, then $\{a_N\} = o(N^\lambda)$ for any $\lambda > 0$.

$(\lim N^{-\lambda}a_N = \lim \frac{\log(N)}{N^\lambda} = \lim \frac{1}{\lambda N^\lambda} = 0)$

(2) $a_N = 10 + \sqrt{N}$, then $a_N = O(N^{1/2})$ and $a_N = o(N^{1/2+\gamma})$ for any $\gamma > 0$.

$(N^{-1/2-\gamma}a_N = \frac{10 + N^{1/2}}{N^{1/2+\gamma}} = \frac{10}{N^{1/2+\gamma}} + \frac{1}{N^\gamma})$

3.2 Convergence in Probability and Bounded in Probability

- **Is your estimator consistent?**
  - consistent: the estimate (estimated parameter) converges to the true value when the sample size goes to infinity.
- **What does the "convergence" mean?**
  - Estimates are calculated with estimator (way of estimation) and data
  - Estimates are random variable because data are random variable (random sample)
  - => convergence in probability !

Convergence in probability

Definition 3.3:

(1) A sequence of random variables $\{x_N : N = 1,2,\cdots\}$ **converges in probability** to the constant a if for all $\varepsilon > 0$, $P[|x_N - a| > \varepsilon] \to 0$ as $N \to \infty$. We write $x_N \xrightarrow{p} a$ and say that a is the plim of x_N.

(2) When $a = 0$, we say that $x_N = o_p(1)$.

(3) A sequence $\{x_N : N = 1,2,\cdots\}$ is **bounded in probability** if and only if for every $\varepsilon > 0$, there exists a $b_\varepsilon < \infty$ and an integer $N_\varepsilon < \infty$ such that $P[|x_N| > b_\varepsilon] < \varepsilon$ for all $N \ge N_\varepsilon$. We write $x_N = O_p(1)$.

Lemma 3.1:

If $x_N \xrightarrow{p} a$, then $x_N = O_p(1)$.

Convergence in probability

Definition 3.4:

(1) A random sequence $\{x_N\}$ is $o_p(a_N)$ where a_N is a nonrandom, positive sequence, if $x_N / a_N = o_p(1)$. We also write $x_N = o_p(a_N)$.

(2) A random sequence $\{x_N\}$ is $O_p(a_N)$ where a_N is a nonrandom, positive sequence, if $x_N / a_N = O_p(1)$. We also write $x_N = O_p(a_N)$.

Example 3.3:

If z is a random variable, $x_N = \sqrt{N} z$ is $O_p(a_N = N^{1/2})$ and $x_N = o_p(a_N = N^{\delta})$ where $\delta > 1/2$.

Useful lemmata

Lemma 3.2:

If $w_N = o_p(1)$, $x_N = o_p(1)$, $y_N = O_p(1)$, and $z_N = O_p(1)$, then

(1) $w_N + x_N = o_p(1)$; (2) $y_N + z_N = O_p(1)$; (3) $y_N z_N = O_p(1)$; (4) $x_N z_N = o_p(1)$.

Lemma 3.4 (Slutsky's theorem):

Let $g : R^K \to R^J$ be a function continuous at some point $c \in R^K$. Let $\{ X_N : N = 1,2,\cdots \}$ be a sequence of K x 1 random vectors such that $x_N \xrightarrow{p} c$. Then $g(x_N) \xrightarrow{p} g(c)$ as $N \to \infty$.

If $g : R^K \to R^J$ is continuous at $\text{plim}\, x_N$, then $p\lim g(x_N) = g(p\lim x_N)$.

3.3 Convergence in Distribution

- What is the asymptotic distribution of your estimator?
  - •Asymptotic distribution is used for statistical tests (ex. Statistical significance of the estimates)
  - •Each estimate (for each sample size N) has a distribution (because it is random)
  - •Which asymptotic distribution does the distribution of the estimate converge to when the sample size N goes to infinity?

- **What does the "convergence" mean?**
  - Convergence in probability: convergence of a random variable to a constant
  - Estimates are random variable which have a distribution
    => convergence in distribution !

Convergence in Distribution

Definition 3.6 :

A sequence of random varibles $\{x_N : N = 1,2,\cdots\}$ **converges in distribution** to the continuous random variable x if and only if $F_N(\xi) \to F(\xi)$ as $N \to \infty$ for all $\xi \in R$ where F_N is the c.d.f. of x_N and F is the continuous c.d.f. of x.

We write $x_N \xrightarrow{d} x$.

If F is normal distribution, we say that "x_N is asymptotically normal."

Definition 3.7 (vector)

Distribution version of the Slutsky's Theorem

Lemma 3.6 (continuous mapping theorem) **:**

Let $\{x_N\}$ be a sequence of K x 1 random vectors such that $x_N \xrightarrow{d} x$.

If $g : R^K \to R^J$ is a continuous function, then $g(x_N) \xrightarrow{d} g(x)$

Corollary 3.2 :

If $\{\mathbf{z}_N\}$ is a sequence of K x 1 random vectors such that $\mathbf{z}_N \xrightarrow{d} \text{Normal}(\mathbf{0}, \mathbf{V})$, then;

(1) For any K x M nonrandom matrix $\mathbf{A}$, $\mathbf{A}'\mathbf{z}_N \xrightarrow{d} \text{Normal}(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A})$.

(2) $\mathbf{z}_N'\mathbf{V}^{-1}\mathbf{z}_N \xrightarrow{d} \chi^2_K$.

(1) Multiplying a constant c on a random variable with the variance sigma^2 results in the distribution with the variance c^2 times sigma^2

(2) The square of a standard normal random variable follows Chi-squared distribution

Asymptotic Equivalence Lemma

Lemma 3.7 :

Let $\{x_N\}$ and $\{z_N\}$ be sequences of K x 1 random vectors.

If $z_N \xrightarrow{d} z$ and $x_N - z_N \xrightarrow{p} 0$, then $x_N \xrightarrow{d} z$

We use this lemma for a comparison of different estimators

Say that we know the asymptotic distribution of one estimator (say z).
If the new estimator (call it x) "converge in probability" to the old estimator (z), then the new estimator follows the same asymptotic distribution as of z in the limit

3.4 Limit Theorems for Random Sampling : weak law of large numbers

Theorem 3.1 (Weak Law of Large Number):

Let $\{w_i : i = 1,2,\cdots\}$ be a sequence of i.i.d. G x 1 random vectors such that $E[|w_{ig}|] < \infty$.

Then the sequence satisfies the weak law of large number (WLLN):

$$N^{-1}\sum_{i=1}^{N} w_i \xrightarrow{p} \mu_w \equiv E(w_i).$$

The sample mean of i.i.d. draws converges in probability to the population mean

WLLN is used to establish consistency of estimators

3.4 Limit Theorems for Random Sampling : Central Limit Theorem

Theorem 3.2 (Lindeberg-Levy):

Let $\{ w_i : i = 1,2,\cdots \}$ be a sequence of i.i.d. G x 1 random vectors such that $E[|w_{ig}|] < \infty$ and $E[w_i] = 0$.

Then $\{ w_i : i = 1,2,\cdots \}$ satisfies the central limit theorem (CLT):

that is, $N^{-1/2}\sum_{i=1}^{N} w_i \xrightarrow{d}$ Normal$(0, \mathrm{B})$, where $\mathrm{B} = Var(w_i) = E(w_i w_i')$ is necessarily positive definite.

The sample mean multiplied by the square-root of N converges in distribution to a normal distribution

CLT guarantees that we can do statistical tests based on a normal distribution when the sample size is large “enough”

See Gallant for the proof using the moment generating function

3.5 Limiting Behavior of Estimators and Test Stat.: 3.5.1 Asymptotic Properties of Estimators

Definition 3.8 (consistency):

Let $\{\hat{\theta}_N : N = 1,2,\cdots\}$ be a sequence of estimators of the P x 1 vector $\theta \in \Theta$,

where N indexes the sample size. If $\hat{\theta}_N \xrightarrow{p} \theta$ for any value of $\theta \in \Theta$,

then we say that $\hat{\theta}_N$ is a consistent estimator of $\theta \in \Theta$.

Since the true $\theta \in \Theta$ is unknown when we estimate it,

consistency requires $\hat{\theta}_N \xrightarrow{p} \theta$ for any $\theta \in \Theta$

Asymptotic Normality

Definition 3.9 (asymptotic normality):

Let $\{\hat{\theta}_N : N = 1,2,\cdots\}$ be a sequence of estimators of the P x 1 vector $\theta \in \Theta$.

Suppose that $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d}$ Normal$(0, \mathbf{V})$ where $\mathbf{V}$ is a P x P positive definite matrix.

Then we say that $\hat{\theta}_N$ is $\sqrt{N}$ asymptotically normally distributed

and $\mathbf{V}$ is the asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta)$, denote Avar$\sqrt{N}(\hat{\theta}_N - \theta) = \mathbf{V}$.

When $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d}$ Normal$(0, \mathrm{V})$ and N is large enough,

we can think as if $\hat{\theta}_N \sim$ Normal$(\theta, \mathrm{V}/N)$ and Avar$(\hat{\theta}_N) = \mathrm{V}/N$

Since V is unknown when we estimate, we replace V by a consistent estimate $\hat{\mathrm{V}}_N$

in the actual estimation: $A\hat{var}(\hat{\theta}_N) = \hat{\mathrm{V}}_N / N$

Asymptotic standard error

Definition 3.10 (asymptotic standard error):

If $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \text{Normal}(0, \mathbf{V})$ where $\mathbf{V}$ is positive definite with jth diagonal v_{jj},
and $\hat{\mathbf{V}}_N \xrightarrow{p} \mathbf{V}$, then the asymptotic standard error of $\hat{\theta}_{Nj}$, denoted $se(\hat{\theta}_{Nj})$, is $(\hat{v}_{Njj}/N)^{1/2}$.

Asymptotic standard error of an estimator is the square root of the appropriate diagonal element of $\hat{\mathbf{V}}_N / N$

$\sqrt{N}$ – **consistent estimator**: $\hat{\theta}_N - \theta = O_p(N^{-1/2})$

The rate (speed) of convergence is almost the square root of the sample size N

That is, $\hat{\theta}_N - \theta = o_p(N^{-c})$ for any $0 \le c < 1/2$.

Asymptotic equivalence and efficiency

Definition 3.11 (efficiency and equivalence):

Let $\hat{\theta}_N$ and $\tilde{\theta}_N$ be estimators of each $\theta \in \Theta$ satisfying $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \text{Normal}(0, \mathbf{V})$
and $\sqrt{N}(\tilde{\theta}_N - \theta) \xrightarrow{d} \text{Normal}(0, \mathbf{D})$.

(1) $\hat{\theta}_N$ is asymptotically efficient relative to $\tilde{\theta}_N$ if $\mathbf{D} - \mathbf{V}$ is positive semidefinite for all $\theta \in \Theta$;

(2) $\hat{\theta}_N$ and $\tilde{\theta}_N$ are root - N equivalent if $\sqrt{N}(\hat{\theta}_N - \tilde{\theta}_N) = o_p(1)$.

If you have several estimators with the square root N asymptotic normality, you prefer one with the “smallest variance”

Asymptotic properties of test statistics

Definition 3.13 :

(1) The asymptotic size of a testing procedure is defined as the limiting probability of rejecting H0 when it is true. Mathematically, we can write this as $\lim_{N\to\infty} P_N(reject\ H_0 \mid H_0)$.

(2) A test is said to be consistent against the alternative H1 if the null hypothesis is rejected with probability approaching one when H1 is true: $\lim_{N\to\infty} P_N(reject\ H_0 \mid H_1) = 1$.

(1) Type I error: Rejecting H0 when H0 is true
The probability of type-one error is called “significance level”

(2) Type II error: cannot reject H0 even if H0 is not true
The probability of type-two error is zero asymptotically with a consistent test

Test of linear constraints

Lemma 3.8 :

Suppose that $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \text{Normal}(0, \mathbf{V})$, where $\mathbf{V}$ is positive definite.

Then for any nonstochastic Q x P matrix $\mathbf{R}$ ($Q \le P$) with rank($\mathbf{R}$) = Q,

$\sqrt{N}\mathbf{R}(\hat{\theta}_N - \theta) \xrightarrow{d} \text{Normal}(0, \mathbf{RVR'})$ and $[\sqrt{N}R(\hat{\theta}_N - \theta)]'[RVR']^{-1}[\sqrt{N}R(\hat{\theta}_N - \theta)] \xrightarrow{d} \chi^2_Q$.

In addition, if $p\lim \hat{\mathbf{V}}_N = \mathbf{V}$, then

$[\sqrt{N}R(\hat{\theta}_N - \theta)]'[R\hat{V}_N R']^{-1}[\sqrt{N}R(\hat{\theta}_N - \theta)] = (\hat{\theta}_N - \theta)'R'[R(\hat{V}_N/N)R']^{-1}R(\hat{\theta}_N - \theta) \xrightarrow{d} \chi^2_Q$

Example: $H_0 : R\theta = r$, where r is a Q x 1 nonrandom vector

Wald statistic: $W_N \equiv (R\hat{\theta}_N - r)'[R(\hat{V}_N / N)R']^{-1}(R\hat{\theta}_N - r)$

Under H0, $W_N \xrightarrow{d} \chi^2_Q$.

Example of linear constraints test

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + u \quad (P = 4)$$

$$H_0 : \theta_1 = 0, \theta_2 = \theta_3 \ (Q = 2)$$

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$W_N \equiv (\mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r})'[\mathbf{R}(\hat{\mathbf{V}}_N / N)\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r})$$

Under H0, $W_N \xrightarrow{d} \chi^2_Q$

Test of nonlinear constraints（delta method）

Lemma 3.9:

Suppose that $\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \text{Normal}(0, V)$, where V is positive definite.

Let $c : \Theta \to R^Q$ be a continuously differentiable function on the parameter space $\Theta \subset R^P$, where $Q \le P$, and assume that θ is in the interior of the parameter space.

Define $C(\theta) \equiv \nabla_\theta c(\theta)$ as the Q x P Jacobian of c. Then,

$$\sqrt{N}(c(\hat{\theta}_N) - c(\theta)) \xrightarrow{d} \text{Normal}(0, C(\theta)VC(\theta)')$$

and $[\sqrt{N}(c(\hat{\theta}_N) - c(\theta))]'[C(\theta)VC(\theta)']^{-1}[\sqrt{N}(c(\hat{\theta}_N) - c(\theta))] \xrightarrow{d} \chi^2_Q$.

Define $\hat{C}_N \equiv C(\hat{\theta}_N)$. Then $p\lim \hat{C}_N = C(\theta)$. If $p\lim \hat{V}_N = V$, then,

$$[\sqrt{N}(c(\hat{\theta}_N) - c(\theta))]'[\hat{C}_N \hat{V}_N \hat{C}_N']^{-1}[\sqrt{N}(c(\hat{\theta}_N) - c(\theta))] \xrightarrow{d} \chi^2_Q.$$

$$Avar(c(\hat{\theta}_N)) = \hat{C}_N [Avar(\hat{\theta}_N)] \hat{C}_N'$$

Use Jacobian when the constraints are non-linear

Test of nonlinear constraints

$$H_0 : c(\theta) = 0$$

Wald statistics $W_N = c(\hat{\theta}_N)'[\hat{C}_N(\hat{V}_N / N)\hat{C}_N']^{-1} c(\hat{\theta}_N)$

Under H0, $W_N \xrightarrow{d} \chi^2_Q$.

When constraints are nonlinear, linearize the constraints by Taylor expansion and apply Lemma 3.8

Problem Set 2

- 3.2
- 3.5
- 3.7
- 3.8
- 3.9